(19) Europäisch s Patentamt
European Patent Office
Office europé n d s brevets

(11) Publication number: **0 117 058**
A2

(12) # EUROPEAN PATENT APPLICATION

(21) Application number: 84300298.1

(22) Date of filing: 18.01.84

(51) Int. Cl.³: C 12 N 15/00
C 12 P 21/02, C 12 N 5/00
//C12R1/91

(30) Priority: 19.01.83 US 459152

(43) Date of publication of application:
29.08.84 Bulletin 84/35

(84) Designated Contracting States:
AT BE CH DE FR GB IT LI LU NL SE

(71) Applicant: GENENTECH, INC.
460 Point San Bruno Boulevard
South San Francisco California 94080(US)

(72) Inventor: Levinson, Arthur David
3690 Ralston Avenue
Hillsborough California 94010(US)

(72) Inventor: Simonsen, Christian Clinton
1509 Parkview Avenue
San Jose California 95310(US)

(74) Representative: Armitage, Ian Michael et al,
MEWBURN ELLIS & CO. 2/3 Cursitor Street
London EC4A 1BQ(GB)

(54) **Methods for producing mature protein in vertebrate host cells and polycistronic expression vectors therefor.**

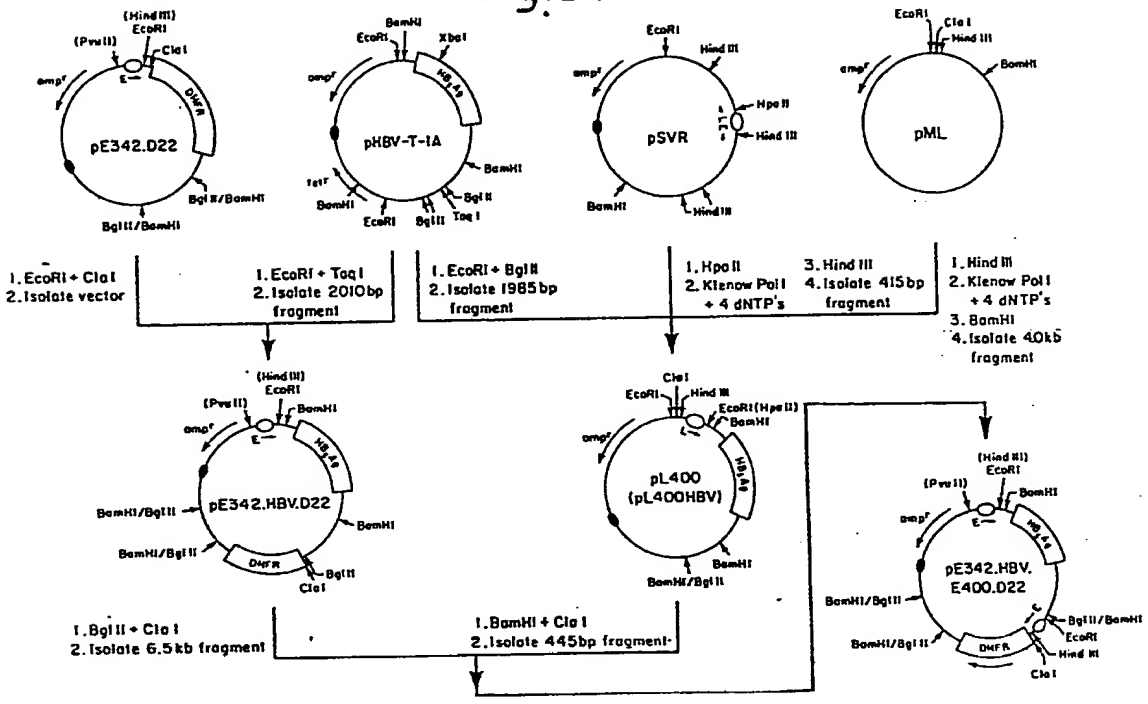
(57) An expression vector capable of expressing in a vertebrate host cell culture a desired protein and a secondary protein, which vector comprises a DNA sequence encoding for a desired protein and a DNA sequence encoding for a secondary protein wherein both said DNA sequences are operably linked to the same promoter sequence and separated by translational stop and start codons.

The secondary sequence provides for a convenient screening marker, both for transformants in general, and for transformants showing high expression levels for the primary sequence, as well as serving as a control device whereby the expression of a desired polypeptide can be regulated, most frequently enhanced.

EP 0 117 058 A2

./...

Fig. 3.


pE342.D22
pHBV-T-1A
pSVR
pML
pE342.HBV.D22
pL400 (pL400HBV)
pE342.HBV. E400.D22
1. EcoRI + Cla I
2. Isolate vector
1. EcoRI + Taq I
2. Isolate 2010bp fragment
1. EcoRI + Bgl II
2. Isolate 1985bp fragment
1. Hpa II
2. Klenow Pol I + 4 dNTP's
3. Hind III
4. Isolate 415bp fragment
1. Hind III
2. Klenow Pol I + 4 dNTP's
3. BamHI
4. Isolate 4.0kb fragment
1. Bgl II + Cla I
2. Isolate 6.5kb fragment
1. BamHI + Cla I
2. Isolate 445bp fragment

Docket 100/92

METHODS FOR PRODUCING MATURE PROTEIN IN VERTEBRATE HOST CELLS AND POLYCISTRONIC EXPRESSION VECTORS THEREFOR

Background of the Invention

This invention relates to the application of recombinant DNA technology to the production of polypeptide in vertebrate cell cultures. More specifically, this invention relates to utilizing the coding sequence for a secondary control polypeptide as a tool in controlling production of a foreign polypeptide by the vertebrate cell culture.

The general principle of utilizing a host cell for the production of a heterologous protein -- i.e., a protein which is ordinarily not produced by this cell -- is well known. However, the technical difficulties of obtaining reasonable quantities of the heterologous protein by employing vertebrate host cells which are desirable by virtue of their properties with regard to handling the protein formed are many. There have been a number of successful examples of incorporating genetic material coding for heterologous proteins into bacteria and obtaining expression thereof. For example, human interferon, desacetyl-thymosin alpha-1, somatostatin, and human growth hormone have been thus produced. Recently, it has been possible to utilize non-bacterial hosts such as yeast cells (see, e.g., co-pending application, U.S. Serial No. 237,913, filed February 25, 1981; EPO Publication No. 0060057) and vertebrate cell cultures (U.S. Application Serial No. 298,235, filed August 31, 1981; EPO Publication No. 0073656) as hosts. The use of vertebrate cell cultures as hosts in the production of mammalian proteins is

advantageous because such systems have additional capabilities for modification, glycosylation, addition of transport sequences, and other subsequent treatment of the resulting peptide produced in the cell. For example, while bacteria may be successfully transfected and caused to express "alpha thymosin", the polypeptide produced lacks the N-acetyl group of the "natural" alpha thymosin found in mammalian system.

In general, the genetic engineering techniques designed to enable host cells to produce heterologous proteins include preparation of an "expression vector" which is a DNA sequence containing,

(1) a "promoter", i.e., a sequence of nucleotides controlling and permitting the expression of a coding sequence;

(2) a sequence providing mRNA with a ribosome binding site;

(3) a "coding region", i.e., a sequence of nucleotides which codes for the desired polypeptide; and

(4) a "termination sequence" which permits transcription to be terminated when the entire code for the desired protein has been read; and

(5) if the vector is not directly inserted into the genome, a "replicon" or origin of replication which permits the entire vector to be reproduced once it is within the cell.

In the construction of vectors in the present invention, the same promoter controls two coding sequences, one for a desired protein, and the other for a secondary protein. Transcription termination is also snared by these sequences. However, the proteins are produced in discrete form because they are separated by a stop and start translational signal.

Ordinarily, the genetic expression vectors are in the form of plasmids, which are extrachromosomal loops of double stranded DNA. These are found in natural form in bacteria, often in multiple copies per cell. However, artificial plasmids can also be

constructed, (and these, of course, are the most useful), by splicing together the four essential elements outlined above in proper sequence using appropriate "restriction enzymes". Restriction enzymes are nucleases whose catalytic activity is limited to lysing at a particular base sequence, each base sequence being characteristic for a particular restriction enzyme. By artful construction of the terminal ends of the elements outlined above (or fractions thereof) restriction enzymes may be found to splice these elements together to form a finished genetic expression vector.

It then remains to induce the host cell to incorporate the vector (transfection), and to grow the host cells in such a way as to effect the synthesis of the polypeptide desired as a concomitant of normal growth.

Two typical problems are associated with the above-outlined procedure. First, it is desirable to have in the vector, in addition to the four essential elements outlined above, a marker which will permit a straightforward selection for those cells which have, in fact, accepted the genetic expression vector. In using bacterial cells as hosts, frequently used markers are resistance to an antibiotic such as tetracycline or ampicillin. Only those cells which are drug resistant will grow in cultures containing the antibiotic. Therefore, if the cell culture which has been sought to be transfected is grown on a medium containing the antibiotic, only the cells actually transfected will appear as colonies. As the frequency of transformation is quite low (approximately 1 cell in 10^6 being transfected under ideal conditions) this is almost an essential prerequisite as a practical matter.

For vertebrate cells as hosts, the transformation rate achieved is more efficient (about 1 cell in 10^3). However, facile selection remains important in obtaining the desired transfected cells. Selection is rendered important, also, because the rate of cell division is about fifty times lower than in bacterial cells --

i.e., although E. coli divide once in about every 20-30 minutes, human tissue culture cells divide only once in every 12 to 24 hours.

The present invention, in one aspect, addresses the problem of selecting for vertebrate cells which have taken up the genetic expression vector for the desired protein by utilizing expression of the coding sequence for a secondary protein, such, for example, as an essential enzyme in which the host cell is deficient. For example, dihydrofolate reductase (DHFR) may be used as a marker using host cells deficient in DHFR.

A second problem attendant on production of polypeptides in a foreign host is recovery of satisfactory quantities of protein. It would be desirable to have some mechanism to regulate, and preferably enhance, the production of the desired heterologous polypeptide. In a second aspect of the invention, a secondary coding sequence which can be affected by externally controlled parameters is utilized to allow control of expression by control of these parameters. Furthermore, provision of both sequences on a polycistron in itself permits selection of transformants with high expression levels of the primary sequence.

It has been shown that DHFR coding sequences can be introduced into, expressed in, and amplified in mammalian cells. Genomic DNA from methotrexate resistant Chinese Hamster Ovary (CHO) cells has been introduced into mouse cells and results in transformants which are also resistant to methotrexate (1). The mechanism by which methotrexate (MTX) resistance in mouse cells is developed appears to be threefold: through gene amplification of the DHFR coding sequence (2,3,4); through decrease in uptake of MTX (5,6) and through reduction in affinity of the DHFR produced for MTX (7).

It appears that amplification of the DHFR gene through MTX exposure can result in a concommitant amplification of a co-transfected gene sequence. It has also been shown that mouse fibroblasts, transfected with both a plasmid containing hepatitis B DNA sequences, and genomic DNA from a hamster cell line containing a mutant gene for MTX-resistant DHFR, secrete increased amounts of hepatitis B surface antigen (HBsAg) into the medium when MTX is employed to stimulate DHFR sequence amplification (8). Further, mRNA coding for the E. coli protein XGPRT is amplified in the presence of MTX in CHO cells co-transfected with the DHFR and XGPRT gene sequences under control by independent promoters (9). Finally, increased expression of a sequence endogenous to the promoter in a DHFR/SV40 plasmid combination in the presence of MTX has been demonstrated (10).

Summary of the Invention

The present invention is based on the discovery that, in vertebrate cell hosts, where the genetic expression vector for a desired polypeptide contains a secondary genetic coding sequence under the control of the same promoter, this secondary sequence provides for a convenient screening marker, both for transformants in general, and for transformants showing high expression levels for the primary sequence, as well as serving as a control device whereby the expression of a desired polypeptide can be regulated, most frequently enhanced.

This is particularly significant as the two proteins, according to the method of this invention, are produced separately in mature form. While both DNA coding sequences are controlled by the same transcriptional promoter, so that a fused message (mRNA) is formed, they are separated by a translational stop signal for the first and start signal for the second, so that two independent proteins result.

As a vertebrate host cell culture system is often advantageous because it is capable of glycosylation, phosphorylation, and lipid association appropriate to animal systems, (whereas bacterial hosts are not), it is significant that marker systems and regulating systems can be provided within this context.

Accordingly, one aspect of the invention is a method for obtaining useful heterologous proteins from vertebrate cell host cultures through the use of a polycistronic expression vector which contains sequences coding for a secondary protein and a desired protein, wherein both the desired and secondary sequences are governed by the same promoter. The coding sequences are separated by translational stop and start signal codons. The expression of the secondary sequence effects control over the expression of the sequence for the desired protein, and the secondary protein functions as a marker for selection of transfected cells. The invention includes use of secondary sequences having either or both of these effects.

In other aspects, the invention concerns the genetic expression vectors suitable for transfecting vertebrate cells in order to produce the desired heterologous peptide, the cell culture produced by this transfection, and the polypeptide produced by this cell culture.

Brief Description of the Drawings

Figure 1 shows the construction of an expression vector for HBsAg, pE342.HS94.HBV.

Figure 2 shows the construction of an expression vector for DHFR, pE342.D22.

Figure 3 shows the construction of an expression vectors for DHFR and HBsAg, pE342.HBV.D22 and pE342.HBV.E400.D22.

Detailed Description and Description of the Preferred Embodiment

A. Definitions

As used herein,

"Plasmids" includes both naturally occurring plasmids in bacteria, and artificially constructed circular DNA fragments.

"Expression vector" means a plasmid which contains at least the four essential elements set forth hereinabove for the expression of the heterologous peptide in a host cell culture.

"Heterologous protein" means a protein or peptide which is not normally produced by, or required for the viability of, the host organism.

"Desired protein" means a heterologous protein or peptide which the method of the invention is designed to produce.

"Secondary peptide" means the protein or peptide which is not the heterologous peptide desired as the primary product of the expression in the host cell, but rather a different heterologous peptide which, by virtue of its own characteristics, or by virtue of the characteristics of the sequence coding for it is capable of "marking" transfection by the expression vector and/or regulating the expression of the primarily desired heterologous peptide.

The peptide sequence may be either long or short ranging from about 5 amino acids to about 1000 amino acids. The conventional distinction between the words peptide and protein is not routinely observed in the description of the invention. If the distinction is to be made, it will be so specified.

"Primary sequence" is the nucleotide sequence coding for the desired peptide, and

"Secondary Sequence" means a sequence of nucleotides which codes for the secondary peptide.

"Transfection" of a host cell means that an expression vector has been taken up by the host cell in a detectable manner whether or not any coding sequences are in fact expressed. In the context of the present invention, successful transfection will be recognized when any indication of the operation of this vector within the host cell is realized. It is recognized that there are various levels of success within its context. First, the vector's coding sequence may or may not be expressed. If the vector is properly constructed with inclusion of promoter and terminator, however, it is highly probable that expression will occur. Second, if the plasmid representing the vector is taken up by the cell and expressed, but fails to be incorporated within the normal chromosomal material of the cell, the ability to express this plasmid will be lost after a few generations. On the other hand, if the vector is taken up within the chromosome, the expression remains stable through repeated replications of the host cell. There may also be an intermediate result. The precise details of the manner in which transfection can thus occur are not understood, but it is clear that a continuum of outcomes is found experimentally in terms of the stability of the expression over several generations of the host culture.

B. A Preferred Embodiment of the Desired Peptide

In a preferred specific embodiment, exemplary of the invention herein, the primary genetic sequence encodes the hepatitis B-surface antigen (HBsAg). This protein is derived from hepatitis B virus, the infective agent of hepatitis B in human beings. This disease is characterized by debilitation, liver damage, primary carcinoma, and often death. The disease is reasonably widespread especially in many African and Asian countries, where many people are chronic carriers with the potential of transmitting the disease pandemically. The virus (HBV) consists of a DNA molecule surrounded

by a nuclear capsid, in turn surrounded by an envelope. Proteins which are associated with the virus include the surface antigen (HBsAg), a core antigen, and a DNA polymerase. The HBsAg is known to produce antibodies in infected people. HBsAg found in the serum of infected individuals consists of protein particles which average ca. 22 nanometers in diameter, and are thus called "22 nanometer particles". Accordingly, it is believed that the HBsAg particle would be an effective basis for a vaccine.

C. A Preferred Embodiment of the Secondary Peptide

It has been recognized that environmental conditions are often effective in controlling the quantity of particular enzymes that are produced by cells under certain growth conditions. In the preferred embodiment of the present invention, advantage is taken of the sensitivity of certain cells to methotrexate (MTX) which is an inhibitor of dihydrofolate reductase (DHFR). DHFR is an enzyme which is required, indirectly, in synthesis reactions involving the transfer of one carbon units. Lack of DHFR activity results in inability of cells to grow except in the presence of those compounds which otherwise require transfer of one carbon units for their synthesis. Cells lacking DHFR, however, will grow in the presence of a combination of glycine, thymidine and hypoxanthine.

Cells which normally produce DHFR are known to be inhibited by methotrexate. Most of the time, addition of appropriate amounts of methotrexate to normal cells will result in the death of the cells. However, certain cells appear to survive the methotrexate treatment by making increased amounts of DHFR, thus exceeding the capacity of the methotrexate to inhibit this enzyme (2,3,4). It has been shown previously that in such cells, there is an increased amount of messenger RNA coding for the DHFR sequence. This is explained by assuming an increase in the amount of DNA in the genetic material coding for this messenger RNA. In effect, apparently the addition of methotrexate causes gene amplification of the DHFR gene. Genetic sequences which are physically connected

with the DHFR sequence although not regulated by the same promoter are also amplified (1,8,9,10). Consequently, it is possible to use the amplification of the DHFR gene resulting from methotrexate treatment to amplify concomitantly the gene for another protein, in this case, the desired peptide.

Moreover, if the host cells into which the secondary sequence for DHFR is introduced are themselves DHFR deficient, DHFR also serves as a convenient marker for selection of cells successfully transfected. If the DHFR sequence is effectively connected to the sequence for the desired peptide, this ability serves as a marker for successful transfection with the desired sequence as well.

D. Vector Construction Techniques Employed (Materials and Methods)

The vectors constructed in the Examples set forth in E are constructed by cleavage and ligation of isolated plasmids or DNA fragments.

Cleavage is performed by treating with restriction enzyme (or enyzmes) in suitable buffer. In general, about 20 µg plasmid or DNA fragments require about 1-5 units of enzyme in 200 µl of buffer solution. (Appropriate buffers for particular restriction enzymes are specified by the manufacturer.) Incubation times of about 1 hour at 37°C are workable. After incubations, protein is removed by extraction with phenol and chloroform, and the nucleic acid recovered from the aqueous fraction by precipitation with ethanol.

If blunt ends are required, the preparation is treated for 15 minutes at 15° with 10 units of Polymerase I (Klenow), phenol-chlorform extracted, and ethanol precipitated.

Size separation of the cleaved fragments is performed using 6 percent polyacrylamide gel described by Goeddel, D. et al., Nucleic Acids Res 8:4057 (1980) incorporated herein by reference.

For ligating approximately equimolar amounts of the desired components, suitably end tailored to provide correct matching are treated with about 10 units T4 DNA ligase per 0.5 µg DNA.

E. Detailed Description of a Preferred Embodiment:

In general, the expression vector suitable for the present invention is constructed by adaptation of gene splicing techniques. The starting material is a naturally occurring bacterial plasmid, previously modified, if desired. A preferred embodiment of the present invention utilizes a pML plasmid which is a modified pBR 322 plasmid prepared according to Lusky, M. et al., Nature 239:79 (1981) which is provided with a single promoter, derived from the simian virus SV-40 and the coding sequence for DHFR and for HBsAg.

In the construction, the promoter (as well as a ribosome binding sequence) is placed upstream from the coding sequence coding for a desired protein and one coding for a secondary protein. A single transcription termination sequence is downstream from both. At the end of the upstream code sequence is placed a translational stop signal; a translational start signal begins the downstream sequence. Thus, expression of the two coding sequences results in a single mRNA strand, but two separate mature proteins.

In a particularly preferred embodiment, the sequence coding for the secondary peptide is downstream from that coding for the desired peptide. Under these circumstances, procedures designed to select for the cells transformed by the secondary peptide will also select for particularly enhanced production of the desired peptide.

F. Examples

The following examples are intended to illustrate, but not limit the invention.

Example 1

Vector Containing the HBsAg Sequence, pE342.HS94.HBV

Figure 1 shows the construction of the HBsAg plasmid.

The 1986 bp EcoRI-BglII fragment which spans the surface antigen gene was isolated from the HBV viral genome cloned with pBR322 as described by Liu et al., DNA 1:213 (1982), incorporated herein by reference. This sequence was ligated between the EcoRI and BamHI sites of pML, a pBR322 derivative which lacks sequences inhibitory to its replication in simian cells, as described by Lusky et al., Nature 293:79 (1981), incorporated herein by reference. Into the single EcoRI site of the resulting plasmid was inserted the 342 bp origin fragment of SV40 obtained by HindIII PvuII digestion of the virus genome, which had been modified to be bounded by EcoRI restriction sites resulting in p342E (also referred to as pHBs348-E) as described by Levinson et al., patent application Serial No. 326,980, filed December 3, 1981, which is hereby incorporated by reference (EPO Publication No. 0073656). (Briefly, the origin of the Simian virus SV40 was isolated by digesting SV40 DNA with HindIII, and converting the HindIII ends to EcoRI ends by the addition of a converter (AGCTGAATTC). This DNA was cut with PvuII, and RI linkers added. Following digestion with EcoRI, the 348 base-pair fragment spanning the origin was isolated by polyacrylamide gel electrophoresis and electroelution, and cloned in pBR322. Expression plasmid pHBs348-E was constructed by cloning the 1986 base-pair fragment resulting from EcoRI and BglII digestion of HBV (Animal Virus Genetics, (Ch. 5) Acad. Press, N.Y. (1980) (which spans the gene encoding HBsAg) into the plasmid pML (Lusky et al., Nature 293:79, 1981) at the EcoRI and BamHI sites. (pML is a derivative of pBR322 which has deletion eliminating sequences which are inhibitory to plasmid replication in monkey cells.) The resulting plasmid (pRI-Bgl) was then linearized with EcoRI, and the 348 base-pair fragment representing the SV40 origin region was introduced into the EcoRI site of pRI-Bgl. The origin fragment can

insert in either orientation. Since this fragment encodes both the early and late SV40 promoters in addition to the origin of replication, HBV genes could be expressed under the control of either promoter depending on this orientation (pHBS348-E representing HBs expressed under control of the early promoter). pE342 is modified by partially digesting with EcoRI, filling in the cleaved site using Klenow DNA polymerase I, and ligating the plasmid back together, thus removing the EcoRI site preceding the SV40 origin in pE342. The resulting plasmid, designated pE342ΔR1, is digested with EcoRI, filled in using Klenow DNA polymerase I, and subcut with BamHI. After electrophoresing on acrylamide gel, the approximately 3500 bp fragment is electroeluted, phenol-chloroform extracted, and ethanol precipitated as above.) The 5' nontranslated leader region of HBsAg was removed by treatment with EcoRI and with Xba, and the analogous 150 bp EcoRI-Xba fragment of a hepatitis expression plasmid pHS94 (Liu et al. (supra)) was inserted in its place to create pE342.HS94.HBV.

(As described by Liu, et al. pHS94 contains the translational start codon of the authentic HBsAg gene, but lacks all 5' nontranslated message sequences. The levels of expression of both the authentic EcoRI-BglII and pHS94 derived equivalent under control of the SV40 early promoter as described above are equivalent and are interchangeable without affecting the performance of the plasmid.)

Example 2

Vector Containing the DHFR Sequence, pE342.D22

A plasmid carrying DHFR as the only expressable sequence is pE348.D22, the construction of which shown in Figure 2.

The 1600 bp Pst I insert of the DHFR cDNA plasmid DHFR-11 (Nunberg et al., Cell 19:355, 1980) was treated with the exonuclease Bal31 in order to remove the poly G:C region adjacent to the Pst I

sites, digested with BglII and the resulting fragments of approximately 660 bp isolated from gels. The Bal31-BglII digested cDNA was ligated into a pBR322 plasmid derivative containing a BglII site. (Following digestion of pBR322 with Hind III, the plasmid fragment was filled in using Klenow DNA polymerase in the presence of the four deoxynucleotide triphosphates, and subcut with BglII.) The resulting plasmid, pDHFR-D22, has an EcoRI site situated 29 bp upstream of the fusion site between pBR322 and the 5' end of the DHFR cDNA. The EcoR I-BglII fragment encompassing the coding sequences of the cDNA insert was then excised from pDHFR-D22 and ligated to EcoRI-BamHI digested pE342.HBV (Example 1), creating the DHFR expression plasmid pE342.D22.

Example 3
Vectors Containing Both DHFR and HBsAg Sequences

Two such vectors were constructed, pE342.HBV.D22 containing a polycistron wherein the DHFR gene is downstream from the HBsAg gene, and pE342.HBV.E400.D22, (Fig. 3) in which the genes coding for DHFR and HBsAg are not polycistronic.

A. pE342.HBV.D22 was constructed by ligating the EcoRI-TaqI fragment of cloned HBV DNA (Liu et al. (supra)), to EcoRI-ClaI digested pE342.D22.

B. This plasmid was further modified by fusing an additional SV40 early promoter between the BglII site and the ClaI site of the DHFR insert of pE342.HBV.D22, creating pE342.HBV.E400.D22.

HBV viral DNA contains a single TaqI site 20 bp beyond the BglII site that was used to generate the EcoRI-BglII fragment encompassing the surface antigen gene. Thus, EcoRI and TaqI digestion of cloned HBV viral DNA results in a fragment of ~2000 bp spanning the surface antigen gene, and containing a single BglII site (1985 bp from the EcoRI site (Liu et al. (supra)). (The ends

of DNA fragments TaqI and ClaI generated by digestion are cohesive, and will ligate together).

The ClaI site is regenerated; thus pE342.HBV.D22 contains both a BglII and ClaI site, which are situated immediately in front of the DHFR coding sequences.

An SV40 origin bounded by restriction sites cohesive with the BglII and ClaI sites of pE342.HBV.D22 was constructed by digesting SV40 DNA with HpaII, filling in as described above, and subcutting with HindIII. A 440 bp fragment spanning the origin was isolated. This was ligated, in a tripartite ligation, to the 4000 bp pBR322 fragment generated by HindIII and BamHI digestion, and the 1986 bp fragment spanning the surface antigen gene generated by digesting the cloned HBV viral DNA with EcoRI, filling in with Klenow DNA polymerase 1, subdigesting with BglII, and isolating on an acrylamide gel. Ligation of all three fragments is achievable only by joining of the filled in HpaII with EcoRI, the two HindIII sites with each other and the BglII with BamHI. Thus when the resulting plasmid is restricted with ClaI and BamHI, a 470 bp fragment is obtained which contains the SV40 origin. This fragment is inserted into the ClaI and BglII sites of pE342.HBV.D22, (paragraph A) creating pE342.HBV.E400.D22 (Fig 3).

Example 4
Transfection of Host Cells

The host cells herein are vertebrate cells grown in tissue culture. These cells, as is known in the art, can be maintained as permanent cell lines prepared by successive serial transfers from isolated normal cells. These cell lines are maintained either on a solid support in liquid medium, or by growth in suspensions containing support nutrients.

In the preferred embodiment, CHO cells, which were deficient in DHFR activity are used. These cells are prepared and propagated as described by Urlaub and Chasin, Proc. Natl. Acad. Sci. (USA) 77:4216 (1980), which is incorporated herein by reference.

The cells are transfected with 5 mg of desired vector as prepared above using the method of Graham and Van Der Eb, Virology 52:456 (1978) incorporated herein by reference.

The method insures the interaction of a collection of plasmids with a particular host cell, thereby increasing the probability that if one plasmid is absorbed by a cell, additional plasmids would be absorbed as well. Accordingly, it is practicable to introduce both the primary and secondary coding sequences using separate vectors for each, as well as by using a single vector containing both sequences.

Example 5
Growth of Transfected Cells and Expression of Peptides

The CHO cells which were subjected to transfection as set forth above were first grown for two days in non-selective medium, then the cells were transferred into medium lacking glycine, hypoxanthine, and thymidine, thus selecting for cells which are able to express the plasmid DHFR. After about 1-2 weeks, individual colonies were isolated with cloning rings.

Cells were plated in 60 or 100 mm tissue culture dishes at approximately .5 x 10^6 cells/dish. After 2 days growth, growth medium was changed. HBsAg was assayed 24 hours later by RIA (Ausria II, Abbott). Cells were counted and HBsAg production standardized on a per cell basis. 10-20 random colonies were analyzed in this fashion for each vector employed.

In one example of the practice of the invention, the following results were obtained:

Vector	Transfectional Efficiency of Dhfr⁻ Cells (Colonies/ug/ 10^6 Cells)	HBsAg Production; ng/10^6 Cells/Day (Percent of Colonies in Given Range)					
		0	0-10	10-100	100-500	500-1500	>1500
pE342.D22	935	100	0	0	0	0	0
pE342.HS94	<.2	1	1	1	1	1	1
pE342.D22+pE342.HS94	340	0	50	30	20	0	0
pE342.HBV.D22	20	0	0	0	0	55	45
pE342.HBV.E400.D22	510	0	17	17	58	8	0

The production of surface antigen in several of the highest expressing cell lines has been monitored for greater than 20 passages and is stable. The cells expressing the surface antigen remain attached to the substratum indefinitely and will continue to secrete the large amounts of surface antigen as long as the medium is replenished.

It is clear that the polycistronic gene construction results in isolation of the cells producing the highest levels of HBsAg. 100 percent of colonies transformed with pE342.HBV.D22 produced over 500 ng/10^6 cells/day whereas 92 percent of those transformed with the non-polycistronic plasmid pE342.HBV.E400.D22 produced less than that amount. Only cells from the polycistronic transfection demonstrated production levels of more than 1500 ng/10^6 cells/day.

Example 6
Treatment with Methotrexate

The surface antigen expressing cell lines are inhibited by methotrexate (MTX), a specific inhibitor of DHFR at concentrations greater than 10 nM. Consistent with previous studies on the effects of MTX on tissue culture cells, occasional clones arise which are

resistant to higher concentrations (50nM) of MTX at a frequency of approximately 10^{-5}. However, these clones no longer produce surface antigen despite the amplification of HBV sequences in the MTX resistant clones. Thus, the HBV gene is amplified, though expression falls off in this case. This suggests that further production of surface antigen may be lethal to the cell.

Example 7
Recovery of Desired Peptide

The surface antigen produced is in the form of a particle, analogous to the 22 nm particle observed in the serum of patients infected with the virus. This form of antigen has been shown to be highly immunogenic. When the cells are grown in medium lacking calf serum or other supplements, approximately 10 percent of the protein contained in the medium is surface antigen and this protein can be isolated by methods known in the art. The surface antigen comigrates on SDS-polyacrylamide gels with the 22 nm particle derived protein.

REFERENCES

1. Wigler, M. et al., Proc. Natl. Acad. Sci. 77:3567 (1980)
2. Schimke, Robert T. et al., Science 202:1051 (1978)
3. Biedler, J.L. et al., Cancer Res. 32:153 (1972)
4. Chang, S.E. et al., Cell 7:391 (1976)
5. Fischer, G.A., Biochem Pharmacol. 11:1233 (1962)
6. Sirotnak, F.M. et al., Cancer Res. 28:75 (1968)
7. Flintoff, W.F. et al., Somat. Cell. Genet. 2:245 (1976)
8. Christman, J. et al., Proc. Natl. Acad. Sci. 79:1815 (1982)
9. Ringold, Gordon et al., J. Molec and Appl. Gen. 1:165 (1981)
10. Kaufman, R.F. et al., J. Molec. Biol. 159:601 (1982)
11. Perucho, Manuel et al., Cell 22:309 (1980)

CLAIMS

1. A method for producing a desired mature protein in a vertebrate host cell, which method comprises:

(a) providing an expression vector which vector comprises

i) a DNA sequence which codes for the desired protein, and

ii) a DNA sequence which codes for a secondary protein whose synthesis is subject to environmental control, and wherein each of the sequences of i) and ii) are positioned so as to be under the control of the same promoter and separated by a translational stop signal, and translational start signal;

(b) transfecting a vertebrate host cell culture with the vector described in (a);

(c) allowing the host cell culture to grow under conditions favorable to the production of the secondary protein.

2. The method of claim 1 wherein the secondary protein is DHFR.

3. The method of claim 2 wherein the host cells are deficient in DHFR.

4. The method of claim 2 or claim 3 wherein the transfected host cell culture is grown in the presence of a DHFR inhibitor.

5. The method of claim 4 wherein the inhibitor is methotrexate.

6. The method of any one of claims 1 to 5 wherein the desired protein is HBsAg.

7. The method of any one of claims 1 to 6 wherein the host cells are CHO cells.

8. A method for controlling the production of a desired protein in a host cell, which method comprises:

(a) transfecting said host cells with an expression vector containing the coding sequences for a secondary protein whose expression is subject to environmental control and for the desired protein both sequences operably linked to the same promoter sequence and separated by a translational stop signal and a translational start signal; and

(b) culturing the cells in the presence of an environmental factor or factors which cause amplification of the sequence for the secondary protein.

9. A method for selecting vertebrate cells which have been transfected with an expression vector capable of expressing a desired protein which method comprises:

treating cells with a vector containing coding sequences for both the desired protein and a secondary protein whose presence is required for the growth of the host cells under selective culture conditions, and

growing the cells under the selective culture conditions;

wherein both coding sequences are operably linked to the same promoter sequence and separated by translational stop and start codons.

10. A method for selecting vertebrate cells which produce high levels of a desired heterologous protein, which method comprises:

treating the cells with a vector containing the coding sequences for a secondary protein whose presence serves as a selection marker for the transfected cells downstream from

the coding sequence for the desired protein; and

growing the cells under selective culture conditions;

wherein both coding sequences are operably linked to the same promoter sequence and separated by translational stop and start signals.

11. The method of claim 9 or claim 10 wherein the secondary protein is DHFR.

12. The method of claim 11 wherein the selective growth conditions comprise a medium lacking glycine , hypoxanthine, and thymidine.

13. An expression vector capable of expressing in a vertebrate host cell culture a desired protein and a secondary protein, which vector comprises a DNA sequence encoding for a desired protein and a DNA sequence encoding for a secondary protein wherein both said DNA sequences are operably linked to the same promoter sequence and separated by translational stop and start codons.

14. The expression vector of claim 13 wherein the coding sequence for the secondary protein encodes for DHFR.

15. The expression vector of claim 13 or claim 14 wherein the promoter sequence is the early promoter derived from SV40.

16. The expression vector of claim 15 which is pE348.HBV.D22.

17. Vertebrate cells transformed with the vector of any one of claims 13 to 16.

18. A polycistronic expression vector which contains the coding sequences for a desired heterologous protein and for

a secondary protein, both operably linked to the same promoter, and separated by translational stop and start codons, wherein the sequence coding for the secondary protein is downstream from the sequence coding for the desired protein.

19. The expression vector of claim 18 wherein the coding sequence for the secondary protein encodes for DHFR.

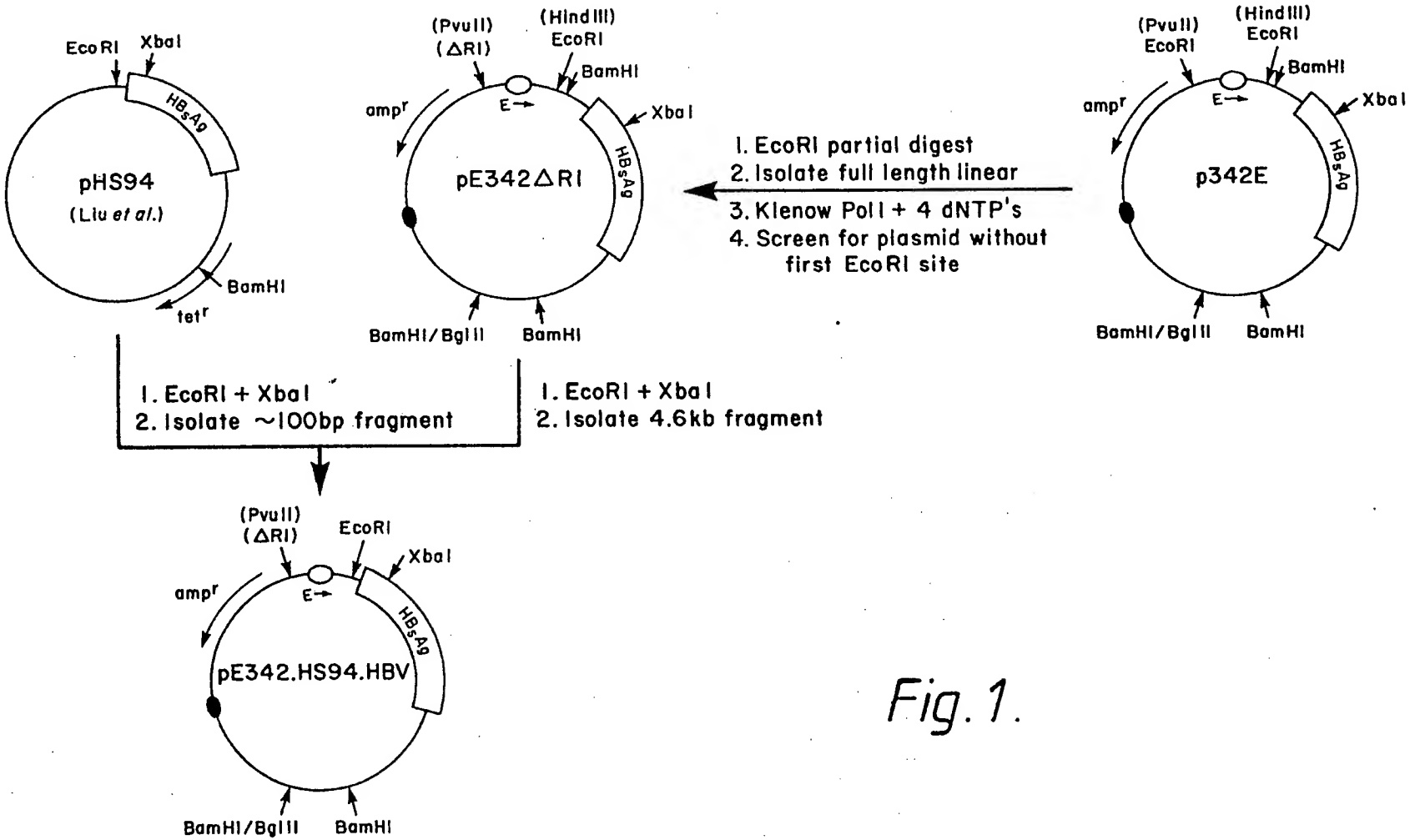


Fig. 1.

0117058

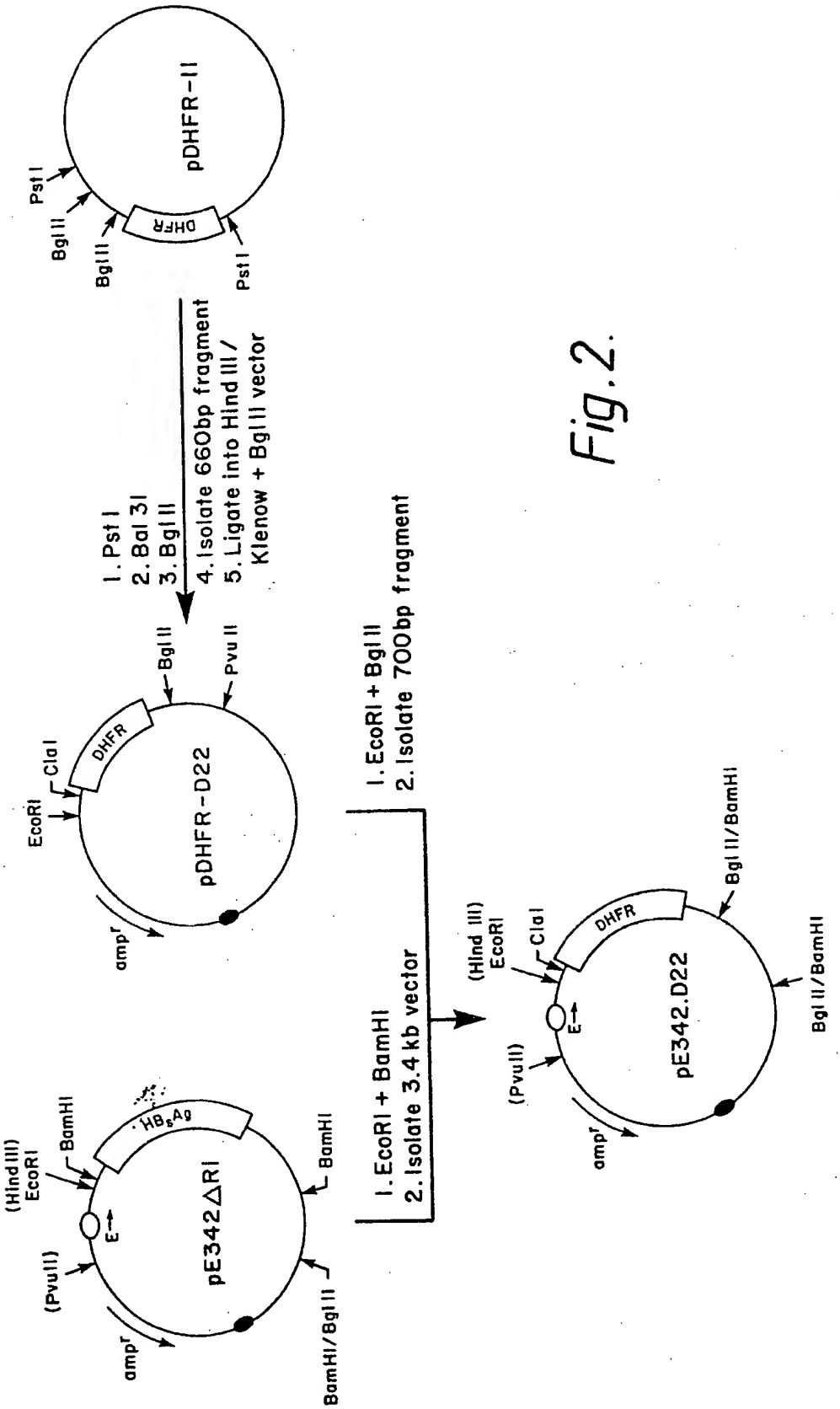
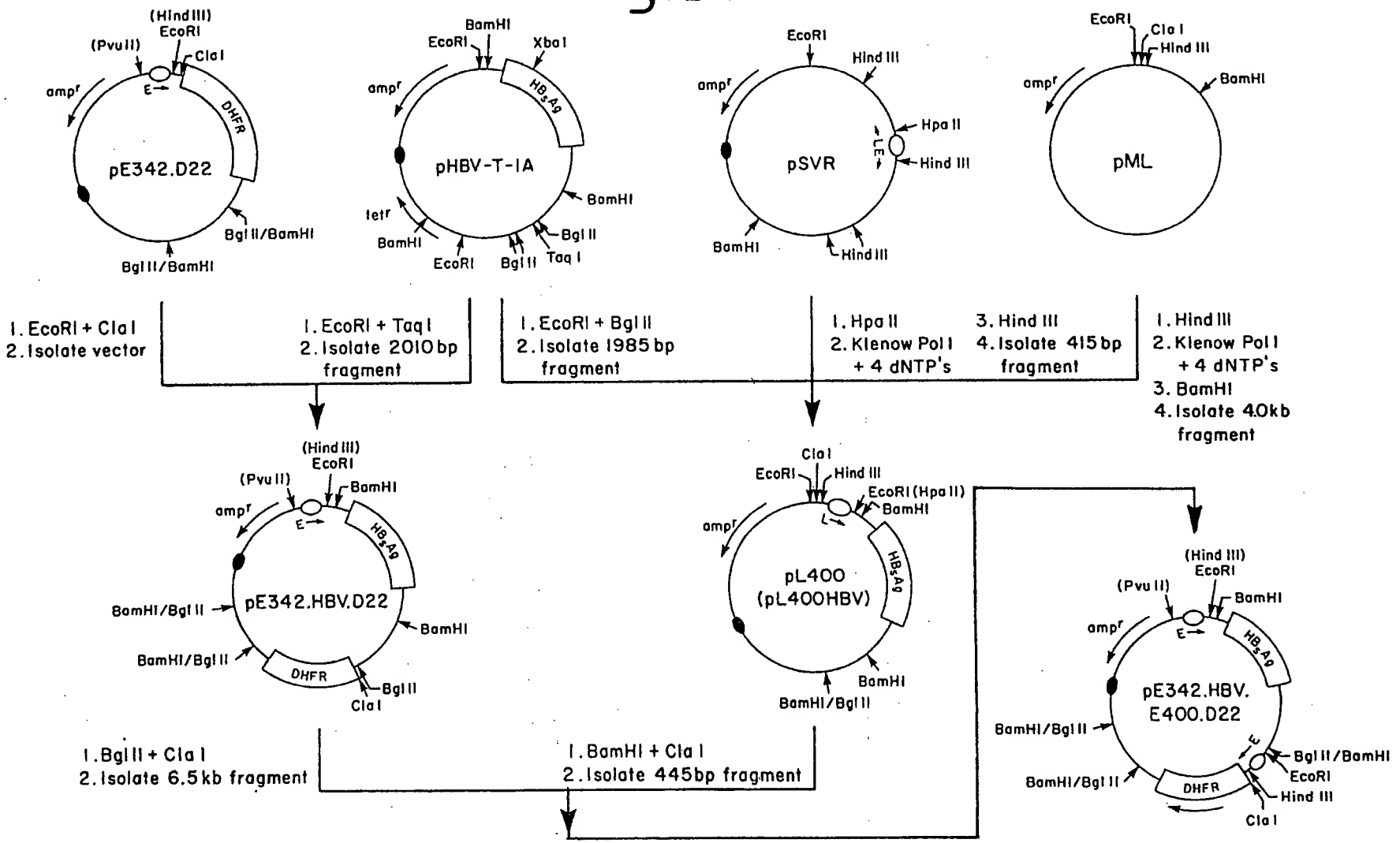

pDHFR-II
Pst I
Bgl II
Bgl II
Pst I
DHFR
1. Pst I
2. Bal 31
3. Bgl II
4. Isolate 660bp fragment
5. Ligate into Hind III / Klenow + Bgl II vector
pDHFR-D22
Bgl II
Pvu II
EcoRI
ClaI
DHFR
ampr
1. EcoRI + Bgl II
2. Isolate 700bp fragment
1. EcoRI + BamHI
2. Isolate 3.4 kb vector
pE342ΔRI
(Hind III)
EcoRI
BamHI
HBsAg
(Pvu II)
E
ampr
BamHI/Bgl II
pE342.D22
(Hind III)
EcoRI
ClaI
DHFR
Bgl II/BamHI
(Pvu II)
E
ampr
Bgl II/BamHI

Fig. 2.

Fig. 3.



3/3

0117058

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 117 058**
B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication of patent specification: **20.09.89**

(51) Int. Cl.⁴: **C 12 N 15/00**, C 12 P 21/02, C 12 N 5/00 // C12R1/91

(21) Application number: **84300298.1**

(22) Date of filing: **18.01.84**

(54) Methods for producing mature protein in vertebrate host cells.

(30) Priority: **19.01.83 US 459152**

(43) Date of publication of application:
29.08.84 Bulletin 84/35

(45) Publication of the grant of the patent:
20.09.89 Bulletin 89/38

(84) Designated Contracting States:
AT BE CH DE FR GB IT LI LU NL SE

(56) References cited:
EP-A-0 009 930
EP-A-0 093 619
EP-A-0 117 059
EP-A-0 117 060
EP-A-0 123 372
WO-A-81/02425
WO-A-81/02425
CHEMICAL ABSTRACTS, vol. 96, no. 21, 24th May 1982, page 174, no. 175313p, Columbus, Ohio, US; G. RINGOLD et al.: "Coexpression and amplification of dihydrofolate reductase cDNA and the Escherichia coli XGPRT gene in Chinese hamster ovary cells" & J. MOL. APPL. GENET. 1981, 1(3), 165-75

(73) Proprietor: **GENENTECH, INC.**
460 Point San Bruno Boulevard
South San Francisco California 94080 (US)

(72) Inventor: **Levinson, Arthur David**
3690 Ralston Avenue
Hillsborough California 94010 (US)
Inventor: **Simonsen, Christian Clinton**
1509 Parkview Avenue
San Jose California 95310 (US)

(74) Representative: **Armitage, Ian Michael et al**
MEWBURN ELLIS & CO. 2/3 Cursitor Street
London EC4A 1BQ (GB)

(56) References cited:
DNA, vol. 1, no. 3, 1st July 1982, pages 213-221, Mary Ann Liebert, Inc., Publishers; CHUNG-CHENG LIU et al.: "Direct expression of hepatitis B surface antigen in monkey cells from an SV40 vector"
EUKARYOTIC VIRAL VECTORS CONFERENCEr, 1982, pages 7-12, Yakov Gluzman, Cold Spring Harbor Lab., Cold Spring Harbor, N.Y., US; D.H. HAMER et al.: "Induction of a mouse metallothionein-I gene in animal virus vectors"

Courier Press, Leamington Spa, England.

EP 0 117 058 B1

(56) References cited:
PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES USA, vol. 77, no. 6, June 1980, pages 3567-3570; M. WIGLER et al.: "Transformation of mammalian cells with an ampliflable dominant-acting gene"
Eukoryotic vital vectors, ed. by Y. Gluzman CSHL 1982, p. 41-45, p. 193-98, p. 145-51
Mol. and Cell Biol. (1981), p. 854-64

Description

Background of the invention

This invention relates to the application of recombinant DNA technology to the production of polypeptide in vertibrate cell cultures. More specifically, this invention relates to utilizing the coding sequence for a secondary control polypeptide as a tool in controlling production of a foreign polypeptide by the vertebrate cell culture.

The general principle of utilizing a host cell for the production of a heterologous protein—i.e., a protein which is ordinarily not produced by this cell—is well known. However, the technical difficulties of obtaining reasonable quantities of the heterologous protein by employing vertebrate host cells which are desirable by virtue of their properties with regard to handling the protein formed are many. There have been a number of successful examples of incorporating genetic material coding for heterologous proteins into bacteria and obtaining expression thereof. For example, human interferon, desacetyl-thymosin alpha-1, somatostatin, and human growth hormone have been thus produced. Recently, it has been possible to utilise non-bacterial hosts such as yeast cells (see, e.g., co-pending application, U.S. Serial No. 237,913, filed February 25, 1981; EPO Publication No. 0060057) and vertebrate cell cultures (U.S. Application Serial No. 298,235, filed August 31, 1981; EPO Publication No. 0073656) as hosts. The use of vertebrate cell cultures as hosts in the production of mammalian proteins is advantageous because such systems have additional capabilities for modification, glycosylation, addition of transport sequences, and other subsequent treatment of the resulting peptide produced in the cell. For example, while bacteria may be successfully transfected and caused to express "alpha thymosin", the polypeptide produced lacks the N-acetyl group of the "natural" alpha thymosin found in mammalian system.

In general, the genetic engineering techniques designed to enable host cells to produce heterologous proteins include preparation of an "expression vector" which is a DNA sequence containing,

(1) a "promoter", i.e., a sequence of nucleotides controlling and permitting the expression of a coding sequence;

(2) a sequence providing mRNA with a ribosome binding site;

(3) a "coding region" i.e., a sequence of nucleotides which codes for the desired polypeptide; and

(4) a "termination sequence" which permits transcription to be terminated when the entire code for the desired protein has been read; and

(5) if the vector is not directly inserted into the genome, a "replicon" or origin of replication which permits the entire vector to be reproduced once it is within the cell.

In the construction of vectors for the method of the present invention, the same promoter controls two coding sequences, one for a desired protein, and the other for a secondary protein. Transcription termination is also shared by these sequences. However, the proteins are produced in discrete form because they are separated by a stop and start translational signal.

Ordinarily, the genetic expression vectors are in the form of plasmids, which are extrachromosomal loops of double stranded DNA. These are found in natural form in baceteria, often in multiple copies per cell. However, artificial plasmids can also be constructed, (and these, of course, are the most useful), by splicing together the four essential elements outlined above in proper sequence using appropriate "restriction enzymes". Restriction enzymes are nucleases whose catalytic activity is limited to lysing at a particular base sequence, each base sequence being characteristic for a particular restriction enzyme. By artful construction of the terminal ends of the elements outlined above (or fractions thereof) restriction enzymes may be found to splice these elements together to form a finished genetic expression vector.

It then remains to induce the host cell to incorporate the vector (transfection), and to grow the host cells in such a way as to effect the synthesis of the polypeptide desired as a concomitant of normal growth.

Two typical problems are associated with the above-outlined procedure. First, it is desirable to have in the vector, in addition to the four essential elements outlined above, a marker which will permit a straightforward selection for those cells which have, in fact, accepted the genetic expression vector. In using bacterial cells as hosts, frequently used markers are resistance to an antibiotic such as tetracycline or ampicillin. Only those cells which are drug resistant will grow in cultures containing the antibiotic. Therefore, if the cell culture which has been sought to be transfected is grown on a medium containing the antibiotic, only the cells actually transfected will appear as colonies. As the frequency of transformation is quite low (approximately 1 cell in 10^6 being transfected under ideal conditions), this is almost an essential prerequisite as a practical matter.

For vertebrate cells as hosts, the transformation rate achieved is more efficient (about 1 cell in 10^3). However, facile selection remains important in obtaining the desired transfected cells. Selection is rendered important, also, because the rate of cell division is about fifty times lower than in baceterial cells—i.e., although *E. coli* divide once in about very 20—30 minutes, human tissue culture cells divide only onc in every 12 to 24 hours.

The present invention, in one aspect, addresses the problem of selecting for vertibrate cells which hav taken up the genetic expr ssion vect r for th desir d pr tein by utilizing expression of th coding sequence for a secondary protein, such, f r exampl , as an ess ntial enzyme in which the host cell is deficient. For example, dihydrofolate reductase (DHFR) may be used as a marker using host cells deficient in DHFR.

A second problem attendant on production of polypeptides in a fermentation host is recovery of satisfactory quantities of protein. It would be desirable to have some mechanism to regulate, and preferably enhance, the production of the desired heterologous polypeptide. In a second aspect of the invention, a secondary coding sequence which can be affected by externally controlled parameters is utilized to allow control of expression by control of these parameters. Furthermore, provision of both sequences on a polycistrom in itself permits selection of transformants with high expression levels of the primary sequence.

It has been shown that DHFR coding sequences can be introduced into, expressed in, and amplified in mammalian cells. Genomic DNA from methotrexate resistant Chinese Hamster Ovary (CHO) cells has been introduced into mouse cells and results in transformants which are also resistant to methotrexate (1). The mechanism by which methotrexate (MTX) resistance in mouse cells is developed appears to be threefold: through gene amplification of the DHFR coding sequence (2, 3, 4); through decrease in uptake of MTX (5, 6) and through reduction in affinity of the DHFR produced for MTX (7).

It appears that amplification of the DHFR gene through MTX exposure can result in a concommitant amplification of a cotransfected gene sequence. It has also been shown that mouse fibroblasts, transfected with both a plasmid containing hepatitis B DNA sequences, and genomic DNA from a hamster cell line containing a mutant gene for MTX-resistant DHFR, secrete increased amounts of hepatitis B surface antigen (HBsAg) into the medium when MTX is employed to stimulate DHFR sequence amplification (8). Further, mRNA coding for the *E. coli* protein XGPRT is amplified in the presence of MTX in CHO cells co-transfected with the DHFR and XGPRT gene sequences under control by independent promoters (9). Finally, increased expression of a sequence endogenous to the promoter in a DHFR/SV40 plasmid combination in the presence of MTX has been demonstrated (10).

It is known that viruses which can infect vertebrate cells often have more than one coding sequence under the control of a single promoter. However, these coding sequences are not expressed simultaneously, but rather they selectively are brought under the control of the promoter for transcription by splicing of the mRNA, for which purpose specific splice sites (donor and acceptor regions) are recognisable in the sequence. Examples of this can be seen in *Eukaryotic Viral Vectors,* ed. Yakov Gluzman, CSHL, 1982, pp 145—151 and pp 193—198.

However, there is no effective intervening splice site, so that translation of the second (downstream) coding sequence would not be expected.

It is also known that in vertebrate cells sometimes translation is not initiated at the first AUG codon in mRNA, but rather from an AUG somewhat downstream, which may be internal to the open reading frame (see for example Subramain et al, *Mol. and Cell. Biol.,* vol. 2, pp 854—864).

The present invention is based on the discovery that, in vertebrate cell hosts, where the genetic expression vector for a desired polypeptide contains a secondary genetic coding sequence under the control of the same promoter but downstream of and separated from the primary coding sequence by translation stop and start codons and without any effective intervening splice site, nevertheless some transfectants may express both sequences, albeit the second more weakly than the first. This secondary sequence can therefore provide for a convenient screening marker, both for transformants in general, and for transformants showing high expression levels for the primary sequence, as well as serving as a control device whereby the expression of a desired polypeptide can be regulated, most frequently enhanced.

This is particularly significant as the two proteins, according to the method of this invention, are produced separately in mature form. While both DNA coding sequences are controlled by the same transcriptional promoter, so that a fused message (mRNA) is formed, they are separated by a translational stop signal for the first and start signal for the second, so that two independent proteins result.

As a vertebrate host cell culture system is often advantageous because it is capable of glycosylation, phosphorylation, and lipid association appropriate to animal systems (whereas bacterial hosts are not), it is significant that marker systems and regulating systems can be provided within this context.

The present invention concerns a method of selecting transfected vertebrate host cells for expression of a desired polypeptide through the use of a polycistronic expression vector which contains sequences coding for a secondary protein and a desired protein, wherein both the desired and secondary sequences are governed by the same promoter. The coding sequences are separated by translational stop and start signal codons. The expression of the secondary sequence effects control over the expression of the sequence for the desired protein, and the secondary protein functions as a marker for selection of transfected cells. The invention includes use of secondary sequences having either or both of these effects.

Brief description of the drawings

Figure 1 shows the construction of an expression vector for HBsAg, pE342.HS94.HBV.

Figure 2 shows the construction of an expression vector for DHFR, pE342.D22.

Figure 3 shows the construction of expression vectors for DHFR and HBsAg, pE342.HBV.D22 and pE342.HBV.E400.D22.

Detailed description and description of the preferred embodiment

A. Definitions

As used herein

"Plasmids" includes both naturally occuring plasmids in bacteria, and artificially constructed circular DNA fragments.

"Expression vector" means a plasmid which contains at least the four essential elements s t forth hereinabove for the expression of the heterologous peptide in a host cell culture.

"Heterologous protein" means a protein or peptide which is not normally produced by, or required for the viability of, the host organism.

"Desired protein" means a heterologous protein or peptide which the method of the invention is designed to produce.

"Secondary peptide" means the protein or peptide which is not the heterologous peptide desired as the primary product of the expression in the host cell, but rather a different heterologous peptide which, by virtue of its own characteristics, or by virtue of the characteristics of the sequence coding for it is capable of "marking" transfection by the expression vector and/or regulating the expression of the primarily desired heterologous peptide.

The peptide sequence may be either long or short ranging from about 5 amino acids to about 1000 amino acids. The conventional distinction between the words peptide and protein is not routinely observed in the description of the invention. If the distinction is to be made, it will be so specified.

"Primary sequence" is the nucleotide sequence coding for the desired peptide, and

"Secondary sequence" means a sequence of nucleotides which codes for the secondary peptide.

"Transfection" of a host cell means that an expression vector has been taken up by the host cell in a detectable manner whether or not any coding sequences are in fact expressed. In the context of the present invention, successful transfection will be recognized when any indication of the operation of this vector within the host cell is realized. It is recognized that there are various levels of success within its context. First, the vector's coding sequence may or may not be expressed. If the vector is properly constructed with inclusion of promoter and terminator, however, it is highly probable that expression will occur. Second, if the plasmid representing the vector is taken up by the cell and expressed, but fails to be incorporated within the normal chromasomal material of the cell, the ability to express this plasmid will be lost after a few generations. On the other hand, if the vector is taken up within the chromosome, the expression remains stable through repeated replications of the host cell. There may also be an intermediate result. The precise details of the manner in which transfection can thus occur are not understood, but it is clear that a continuum of outcomes is found experimentally in terms of the stability of the expression over several generations of the host culture.

B. A preferred embodiment of the desired peptide

In a preferred specific embodiment, exemplary of the invention herein, the primary genetic sequence encodes the hepatitis B-surface antigen (HBsAg). This protein is derived from hepatitis B virus, the infective agent of hepatitis B in human beings. This disease is characterized by debilitation, liver damage, primary carcinoma, and often death. The disease is reasonably widespread especially in many African and Asian countries, where many people are chronic carriers with the potential of transmitting the disease pandemically. The virus (HBV) consists of a DNA molecule surrounded by a nuclear capsid, in turn surrounded by an envelope. Proteins which are associated with the virus include the surface antigen (HBsAg), a core antigen, and a DNA polymerase. The HBsAg is known to produce antibodies in infected people. HBsAg found in the serum of infected individuals consists of protein particles which average ca. 22 nanometers in diameter, and are thus called "22 nanometer particles". Accordingly, it is believed that the HBsAg particle would be an effective basis for a vaccine.

C. A preferred embodiment of the secondary peptide

It has been recognized that environmental conditions are often effective in controlling the quantity of particular enzymes that are produced by cells under certain growth conditions. In the preferred embodiment of the present invention, advantage is taken of the sensitivity of certain cells to methotrexate (MTX) which is an inhibitor of dihydrofolate reductase (DHFR). DHFR is an enzyme which is required, indirectly, in synthesis reactions involving the transfer of one carbon units. Lack of DHFR activity results in inability of cells to grow except in the presence of those compounds which otherwise require transfer of one carbon units for their synthesis. Cells lacking DHFR, however, will grow in the presence of a combination of glycine, thymidine and hypoxanthine.

Cells which normally produce DHFR are known to be inhibited by methotrexate. Most of the time, addition of appropriate amounts of methotrexate to normal cells will result in the death of the cells. However, certain cells appear to survive the methotrexate treatment by making increased amounts of DHFR, thus exceeding the capacity of the methotrexate to inhibit this enzyme (2, 3, 4). It has been shown previously that in such cells, there is an increased amount of messenger RNA coding for the DHFR sequence. This is explained by assuming an increase in the amount of DNA in the genetic material coding for this messenger RNA. In effect, apparently the addition of methotrexat causes g n amplification of the DHFR gene. Genetic sequences which ar physically connected with the DHFR sequence although not regulated by the same promoter are also amplified (1, 8, 9, 10). Consequently, it is possible to use the amplification of th DHFR gene resulting from methotrexate treatment to amplify concomitantly the gene for another protein, in this case, the desired peptide.

Moreov r, if the h st cells into which the s condary sequence for DHFR is introduced are th mselves DHFR deficient, DHFR also serves as a conv nient marker for selection of cells successfully transfected. If

the DHFR sequence is effectively connected to the sequence for the desir d peptide, this ability serves as a marker for successful transfection with the desired sequence as well.

D. Vector construction techniques employed (materials and methods)

The vectors constructed in the Examples set forth in E are constructed by cleavage and ligation of isolated plasmids or DNA fragments.

Cleavage is performed by treating with restriction enzyme (or enzymes) in suitable buffer. In general, about 20 µg plasmid or DNA fragments require about 1—5 units of enzyme in 200 µl of buffer solution. (Appropriate buffers for particular restriction enzymes are specified by the manufacturer). Incubation times of about 1 hour at 37°C are workable. After incubations, protein is removed by extraction with phenol and chloroform, and the nucleic acid recovered from the aqueous fraction by precipitation with ethanol.

If blunt ends are required, the preparation is treated for 15 minutes at 15° with 10 units of Polymerase I (Klenow), phenol-chloroform extracted, and ethanol precipitated.

Size separation of the cleaved fragments is performed using 6 percent polyacrylamide gel described by Goeddel, D. *et al., Nucleic Acids Res 8;* 4057 (1980) incorporated herein by reference.

For ligating approximately equimolar amounts of the desired components, suitably end tailored to provide correct matching are treated with about 10 units T4 DNA ligase per 0.5 µg DNA.

E. Detailed description of a preferred embodiment:

In general, the expression vector suitable for the present invention is constructed by adaptation of gene splicing techniques. The starting material is a naturally occuring bacterial plasmid, previously modified, if desired. A preferred embodiment of the present invention utilizes a pML plasmid which is a modified pBR 322 plasmid prepared according to Lusky, M. *et al., Nature 239:*79 (1981) which is provided with a single promoter, derived from the simian virus SV-40 and the coding sequence for DHFR and for HBsAg.

In the construction, the promoter (as well as a ribosome binding sequence) is placed upstream from the coding sequence coding for a desired protein and one coding for a secondary protein. A single transcription termination sequence is downstream from both. At the end of the upstream code sequence is placed a translational stop signal; a translational start signal begins the downstream sequence. Thus, expression of the two coding sequences results in a single mRNA strand, but two separate mature proteins.

In a particularly preferred embodiment, the sequence coding for the secondary peptide is downstream from that coding for the desired peptide. Under these circumstances, procedures designed to select for the cells transformed by the secondary peptide will also select for particularly enhanced production of the desired peptide.

F. Examples

The following examples are intended to illustrate, but not limit the invention.

Example 1

Vector containing the HBsAg sequence, pE342.HS94.HBV

Figure 1 shows the construction of the HBsAg plasmid.

The 1986 bp EcoRI-BglII fragment which spans the surface antigen gene was isolated from the HBV viral genome cloned with pBR322 as described by Liu *et al., DNA 1:*213 (1982), incorporated herein by reference. This sequence was ligated between the EcoRI and BamHI sites of pML, a pBR322 derivative which lacks sequences inhibitory to its replication in simian cells, as described by Lusky *et al., Nature 293:*79 (1981), incorporated herein by reference. Into the single EcoRI site of the resulting plasmid was inserted the 342 bp origin fragment of SV40 obtained by HindIII PvuII digestion of the virus genome, which had been modified to be bounded by EcoRI restriction sites resulting in p342E (also referred to as pHBs348-E) as described by Levinson *et al.*, patent application Serial No. 326,980, filed December 3, 1981, which is hereby incorporated by reference (EPO Publication No. 0073656). (Briefly, the origin of the Simian virus SV40 was isolated by digesting SV40 DNA with HindIII, and converting the HindIII ends to EcoRI ends by the addition of a converter (AGCTGAATTC). This DNA was cut with PvuII, and RI linkers added. Following digestion with EcoRI, the 348 base-pair fragment spanning the origin was isolated by polyacrylamide gel electrophoresis and electroelution, and cloned in pBR322. Expression plasmid pHBs348-E was constructed by cloning the 1986 base-pair fragment resulting from EcoRI and BglII digestion of HBV (*Animal Virus Genetics,* (CH. 5) Acad. Press, N. Y. (1980) (which spans the gene encoding HBsAg) into the plasmid pML (Lusky *et al., Nature 293*:79, 1981) at the EcoRI and BamHI sites. (pML is a derivative of pBR322 which has deletion eliminating sequences which are inhibitory to plasmid replication in monk y cells). The resulting plasmid (pRI-Bgl) was th n lin arized with Ec RI, and the 348 base-pair fragment representing the SV40 origin region was introduced into the EcoRI site of pRI-Bgl. The origin fragment can insert in either orientation. Since this fragment encodes both th early and late SV40 promoters in addition to th origin of replication, HBV gen s could be expressed under the control f either promoter dep nding on this orientation (pHBS348-E representing HBs expressed under control of the early promoter). pE342 is modified by partially digesting with EcoRI, filling in the cleaved site using Klenow DNA polymerase I, and ligating the plasmid back together, thus removing th EcoRI site preceding the SV40

origin in pE342. The resulting plasmid, designated pE342ΔR1, is digested with EcoRI, filled in using Klenow DNA polymerase I, and subcut with BamHI. After electrophoresing on acrylamide gel, the approximately 3500 bp fragment is electroeluted, phenol-chloroform extracted, and ethanol precipitated as above). The 5' nontranslated leader region of HBsAg was removed by treatment with EcoRI and with Xba, and the analogous 150 bp EcoRI-Xba fragment of a hepatitis expression plasmid pHS94 (Liu *et al. (supra)*) was inserted in its place to create pE342.HS94.HBV.

(As described by Liu, *et al.* pHS94 contains the translational start codon of the authentic HBsAg gene, but lacks all 5' nontranslated message sequences. The levels of expression of both the authentic EcoRI-BglII and pHS94 derived equivalent under control of the SV40 early promoter as described above are equivalent and are interchangeable without affecting the performance of the plasmid).

Example 2

Vector containing the DHFR sequence, pE342.D22

A plasmid carrying DHFR as the only expressable sequence is pE348.D22, the construction of which is shown in Figure 2.

The 1600 bp Pst I insert of the DHFR cDNA plasmid DHFR-11 (Nunberg *et al., Cell 19*:355, 1980) was treated with the exonuclease Bal31 in order to remove the poly G:C region adjacent to the Pst I sites, digested with BglII and the resulting fragments of approximately 660 bp isolated from gels. The Bal31-BglII digested cDNA was ligated into a pBR322 plasmid derivative containing a BglII site. (Following digestion of pBR322 with Hind III, the plasmid fragment was filled in using Klenow DNA polymerase in the presence of the four deoxynucleotide triphosphates, and subcut with BglII). The resulting plasmid, pDHFR-D22, has an EcoRI site situated 29 bp upstream of the fusion site between pBR322 and the 5' end of the DHFR cDNA. The EcoR I-BglII fragment encompassing the coding sequences of the cDNA insert was then excised from pDHFR-D22 and ligated to EcoRI-BamHI digested pE342.HBV (Example 1), creating the DHFR expression plasmid pE342.D22.

Example 3

Vectors containing both DHFR and HBsAg sequences

Two such vectors were constructed, pE342.HBV.D22 containing a polycistron wherein the DHFR gene is downstream from the HBsAg gene, and pE342.HBV.E400.D22, (Figure 3) in which the genes coding for DHFR and HBsAg are not polycistronic.

A. pE342.HBV.D22 was constructed by ligating the EcoRI-TaqI fragment of cloned HBV DNA (Liu *et al. (supra)*, to EcoRI-ClaI digested pE342.D22.

B. This plasmid was further modified by fusing an additional SV40 early promoter between the BglII site and the ClaI site of the DHFR insert of pE342.HBV.D22, creating pE342.HBV.E400.D22.

HBV viral DNA contains a single TaqI site 20 bp beyond the BglII site that was used to generate the EcoRI-BglII fragment encompassing the surface antigen gene. Thus, EcoRI and TaqI digestion of cloned HBV viral DNA results in a fragment of ~2000 bp spanning the surface antigen gene, and containing a single BglII site (1985 bp from the EcoRI site (Liu *et al. (supra)*. (The ends of DNA fragments TaqI and ClaI generated by digestion are cohesive, and will ligate together).

The ClaI site is regenerated; thus pE342.HBV.D22 contains both a BglII and ClaI site, which are situated immediately in front of the DHFR coding sequences.

An SV40 origin bounded by restriction sites cohesive with the BglII and ClaI sites of pE342.HBV.D22 was constructed by digesting SV40 DNA with HpaII, filling in as described above, and subcutting with HindIII. A 440 bp fragment spanning the origin was isolated. This was ligated, in a tripartite ligation, to the 4000 bp pBR322 fragment generated by HindIII and BamHI digestion, and the 1986 bp fragment spanning the surface antigen gene generated by digesting the cloned HBV viral DNA with EcoRI, filling in with Klenow DNA polymerase 1, subdigesting with BglII, and isolating on an acrylamide gel. Ligation of all three fragments is achievable only by joining of the filled in HpaII with EcoRI, the two HindIII sites with each other and the BglII with BamHI. Thus when the resulting plasmid is restricted with ClaI and BamHI, a 470 bp fragment is obtained which contains the SV40 origin. This fragment is inserted into the ClaI and BglII sites of pE342.HBV.D22, (paragraph A) creating pE342.HBV.E400.D22 (Figure 3).

Example 4

Transfection of host cells

The host cells herein are vertebrate cells grown in tissue culture. These cells, as is known in the art, can be maintained as permanent cell lines prepared by successive serial transfers from isolated normal cells. These cell lines are maintained either on a solid support in liquid medium, or by growth in suspensions containing support nutrients.

In the preferred embodiment, CHO cells, which were deficient in DHFR activity are used. These cells are prepared and propagated as described by Urlaub and Chasin, *Proc. Natl. Acad. Sci* (USA) *77*:4216 (1980), which is incorporated herein by reference.

The cells are transfected with 5 mg of desired vector as prepared above using the method of Graham and Van Der Eb, *Virology 52*:456 (1978) incorporated herein by reference.

The method insures the interaction of a collection of plasmids with a particular host cell, thereby

increasing the probability that if one plasmid is absorbed by a cell, additional plasmids would be absorbed as well. Accordingly, it is practicable to introduce both the primary and secondary coding sequences using separate vectors for each, as well as by using a single vector containing both sequences.

Example 5

Growth of transfected cells and expression of peptides

The CHO cells which were subjected to transfection as set forth above were first grown for two days in non-selective medium, then the cells were transferred into medium lacking glycine, hypoxanthine, and thymidine, thus selecting for cells which are able to express the plasmid DHFR. After about 1—2 weeks, individual colonies were isolated with cloning rings.

Cells were plated in 60 or 100 mm tissue culture dishes at approximately $.5 \times 10^6$ cells/dish. After 2 days growth, growth medium was changed. HBsAg was assayed 24 hours later by RIA (Ausria II, Abbott). Cells were counted and HBsAg production standardized on a per cell basis. 10—20 random colonies were analyzed in this fashion for each vector employed.

In one example of the practice of the invention, the following results were obtained:

Vector	Transfectional efficiency of $Dhfr^-$ cells (colonies/ug/ 10^6 cells)	HBsAg production; ng/10^6 cells/day (percent of colonies in given range)					
		0	0—10	10—100	100—500	500—1500	>1500
pE342.D22	935	100	0	0	0	0	0
pE342.HS94	<.2	1	1	1	1	1	1
pE342.D22+pE342.HS94	340	0	50	30	20	0	0
pE342.HBV.D22	20	0	0	0	0	55	45
pE342.HBV.E400.D22	510	0	17	17	58	8	0

The production of surface antigen in several of the highest expressing cell lines has been monitored for greater than 20 passages and is stable. The cells expressing the surface antigen remain attached to the substratum indefinitely and will continue to secrete the large amounts of surface antigen as long as the medium is replenished.

It is clear that the polycistronic gene construction results in isolation of the cells producing the highest levels of HBsAg. 100 percent of colonies transformed with pE342.HBV.D22 produced over 500 ng/10^6 cells/day whereas 92 percent of those transformed with the non-polycistronic plasmid pE342.HBV.E400.D22 produced less than that amount. *Only* cells from the polycistronic transfection demonstrated production levels of more than 1500 ng/10^6 cells/day.

Example 6

Treatment with methotrexate

The surface antigen expressing cell lines are inhibited by methotrexate (MTX), a specific inhibitor of DHFR at concentrations greater than 10 nM. Consistent with previous studies on the effects of MTX on tissue culture cells, occasional clones arise which are resistant to higher concentrations (50 nM) of MTX at a frequency of approximately 10^{-5}. However, these clones no longer produce surface antigen despite the amplification of HBV sequences in the MTX resistant clones. Thus, the HBV gene is amplified, though expression falls off in this case. This suggests that further production of surface antigen may be lethal to the cell.

Example 7

Recovery of desired peptide

The surface antigen produced is in the form of a particle, analogous to the 22 nm particle observed in the serum of patients infected with the virus. This form of antigen has been shown to be highly immunogenic. When the cells are grown in medium lacking calf serum or other supplements, approximately 10 percent of the protein contained in the medium is surface antigen and this protein can be isolated by methods known in the art. The surface antigen comigrates on SDS-polyacrylamide gels with the 22 nm particle derived protein.

References

1. Wigler, M. *et al., Proc. Natl. Acad. Sci. 77*:3567 (1980)
2. Schimke, Robert T. *et al., Science 202*:1051 (1978)
3. Biedl r, J. L. *et al., Cancer Res. 32*:153 (1972)

4. Chang, S. E. *et al., Cell 7*:391 (1976)
5. Fischer, G. A., *Biochem Pharmacol. 11*:1233 (1962)
6. Sirotnak, F. M. *et al., Cancer Res. 28*:75 (1968)
7. Flintoff, W. F. *et al., Somat. Cell. Genet. 2*:245 (1976)
8. Cristman, J. *et al., Proc. Natl. Acad. Sci. 79*:1815 (1982)
9. Ringold, Gordon *et al., J. Molec and Appl. Gen. 1*:165 (1981)
10. Kaufman R. F. *et al., J. Molec. Biol. 159*:601 (1982)
11. Perucho, Manuel *et al., Cell 22*:309 (1980)

Claims

1. A method of selecting transfected vertebrate host cells for expression of a desired polypeptide, which method comprises transfecting the vertebrate cells with an expression vector comprising a promoter operable in a vertebrate host cell and first and second polypeptide coding sequences under the control of said promoter, said coding sequences being separated by a translational stop signal and a translational start signal without any intervening splice site which is effective in the host cell and selecting transfectants which exhibit expression of the second polypeptide accompanied by a higher level of expression of said first polypeptide.

2. A method according to Claim 1, wherein the second polypeptide is capable of marking transfection by the expression vector and/or regulating the expression of the first polypeptide.

3. A method according to Claim 1 or Claim 2 wherein the promoter is the SV40 early promoter.

4. A method according to any one of the preceding claims wherein the transfectants are grown under selective culture conditions favouring expression of the second polypeptide.

5. A method according to any one of the preceding claims wherein the second polypeptide is DHFR.

6. A method according to Claim 5 wherein the host cells are deficient in DHFR.

7. A method according to Claim 5 or Claim 6 wherein the transfected host cell is grown in the presence of a DHFR inhibitor.

8. A method according to Claim 7 wherein the DHFR inhibitor is methotrexate.

9. A method according to any one of the preceding claims wherein the host cells are CHO cells.

10. A method for producing a desired polypeptide which comprises culturing transfected vertebrate host cells obtained according to any one of the preceding claims so as to express said first polypeptide coding sequence, said first polypeptide being the desired polypeptide.

Patentansprüche

1. Verfahren zur Selektion von transfizierten Wirbeltier-Wirtszellen für die Expression eines gewünschten Polypeptides, welches Verfahren das Transfizieren der Wirbeltierzellen mit einem Expressionsvektor, der einen Promoter, der in einer Wirbeltierwirtszelle operabel ist, und erste und zweite Polypeptidkodierungssequenzen unter der Steuerung des genannten Promotors enthält, wobei die genannten Kodierungssequenzen durch ein Translationsstopsignal und ein Translationsstartsignal ohne intervenierende Spleißstelle getrennt sind, und der in der Wirtszelle wirkt, und die Selektion von Transfektanten, die die Expression des zweiten Polypeptides, begleitet von einem höheren Expressionswert des genannten ersten Polypeptides zeigen, umfaßt.

2. Verfahren nach Anspruch 1, worin das zweite Polypeptid fähig ist, die Transfektion durch den Expressionsvektor zu markieren und/oder die Expression des ersten Polypeptides zu steuern.

3. Verfahren nach Anspruch 1 oder 2, worin der Promotor der frühe SV40 Promoter ist.

4. Verfahren nach einem der vorhergehenden Ansprüche, worin die Transfektanten unter selektiven Kulturbedingungen kultiviert werden, die die Expression des zweiten Polypeptides fördern.

5. Verfahren nach einem der vorhergehenden Ansprüche, worin das zweite Polypeptid DHFR ist.

6. Verfahren nach Anspruch 5, worin die Wirtszellen arm an DHFR sind.

7. Verfahren nach Anspruch 5 oder 6, worin die transfizierte Wirtszelle in Gegenwart eines DHFR-Inhibitors kultiviert wird.

8. Verfahren nach Anspruch 7, worin der DHFR Inhibitor Methotrexat ist.

9. Verfahren nach einem der vorhergehenden Ansprüche, worin die Wirtzellen CHO-Zellen sind.

10. Verfahren zur Herstellung eines gewünschten Polypeptides, welches das Kultivieren von transfizierten Wirbeleiter-Wirtszellen umfaßt, die nach einem der vorhergehenden Ansprüche erhalten wurden, um die genannte erste Polypeptid-Kodierungssequenz zu exprimieren, wobei das genannte erste Polypeptid das gewünschte Polypeptid ist.

Revendications

1. Procédé de sélection de cellules hôtes transfectées de vertébrés pour l'expression d'un polypeptide souhaité, lequel procédé comprend la transfection des cellules de vertébrés avec un vecteur d'expression comprenant un promoteur utilisable dans une cellule hôte de vertébré t des première et seconde séquences de codage de polypeptide sous le contrôl dudit promoteur, l sdit s séquences de codage étant

séparées par un signal d'arrêt de traduction et un signal de début de traduction sans aucin site intermédiaire d'épissage qui est efficace dans la cellule hôte et la sélection de transfectants qui présentent l'expression du second polypeptide accompagnée d'un niveau supérieur d'expression dudit premier polypeptide.

2. Procédé selon la revendication 1 où le second polypeptide est capable de marquer la transfection par le vecteur d'expression et/ou de réguler l'expression du premier polypeptide.

3. Procédé selon la revendication 1 ou la revendication 2 où le promoteur est le promoteur précoce de SV40.

4. Procédé selon l'une quelconque des revendications précédentes où les transfectants sont développés en conditions de culture sélective favorisant l'expression du second polypeptide.

5. Procédé selon l'une quelconque des revendications précédentes où le second polypeptide est DHFR.

6. Procédé selon la revendication 5 où les cellules hôtes sont déficientes en DHFR.

7. Procédé selon la revendication 5 ou la revendication 6 où la cellule hôte transfectée est développée en présence d'un inhibiteur de DHFR.

8. Procédé selon la revendication 7 où l'inhibiteur de DHFR est le méthotrexate.

9. Procédé selon l'une quelconque des revendications précédentes où les cellules hôtes sont des cellules de CHO.

10. Procédé de production d'un polypeptide souhaité, qui comprend la mise en culture de cellules hôtes transfectées de vertébrés obtenues selon l'une quelconque des revendications précédentes afin d'exprimer la séquence codant ledit premier polypeptide, ledit premier polypeptide étant le polypeptide souhaité.

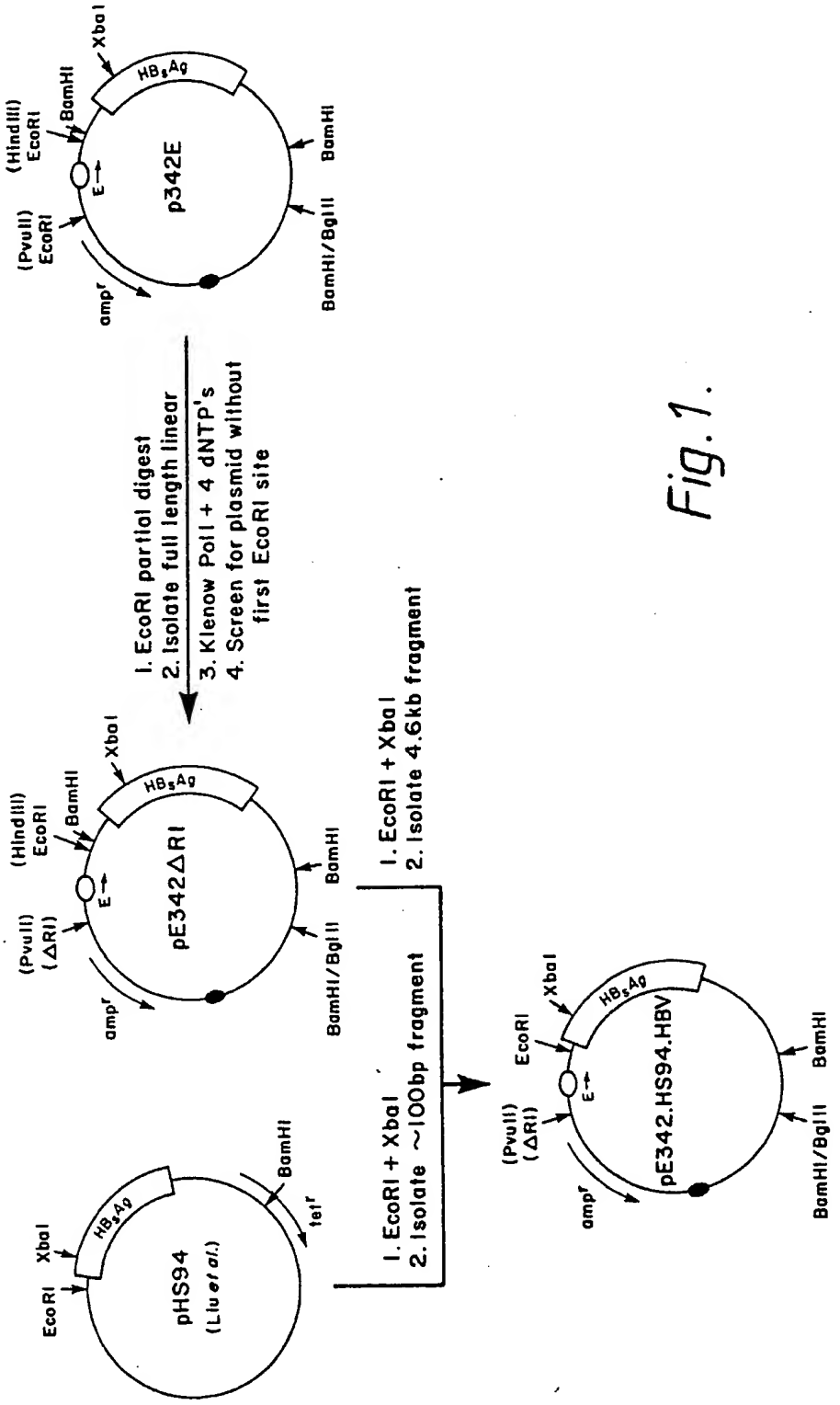

p342E
(HindIII)
EcoRI
BamHI
XbaI
HBsAg
(PvuII)
EcoRI
ampr
BamHI/BglII
BamHI
1. EcoRI partial digest
2. Isolate full length linear
3. Klenow PolI + 4 dNTP's
4. Screen for plasmid without first EcoRI site
pE342ΔRI
(PvuII)
(ΔRI)
1. EcoRI + XbaI
2. Isolate 4.6kb fragment
pHS94
(Liu et al.)
tetr
1. EcoRI + XbaI
2. Isolate ~100bp fragment
pE342.HS94.HBV

Fig. 1.

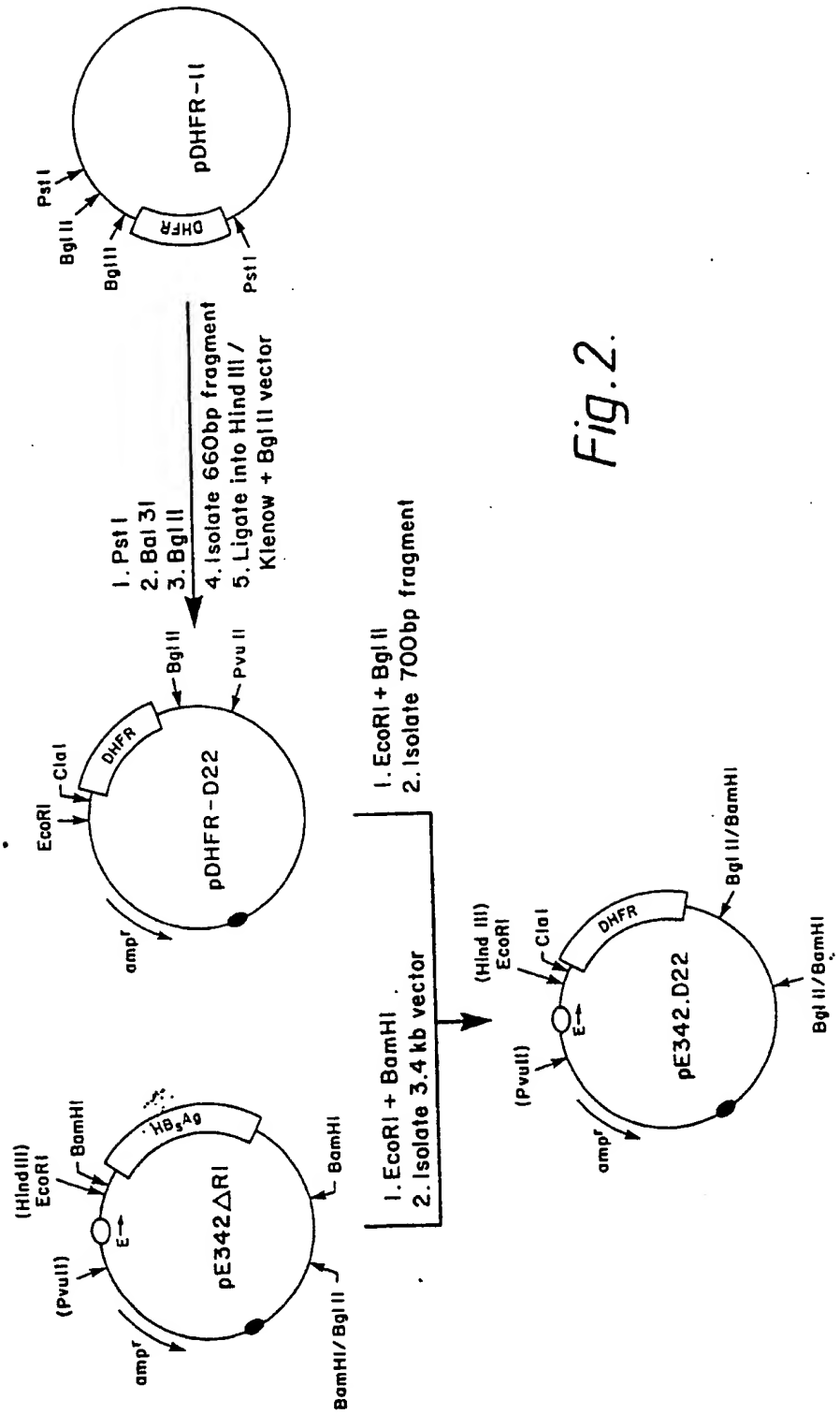
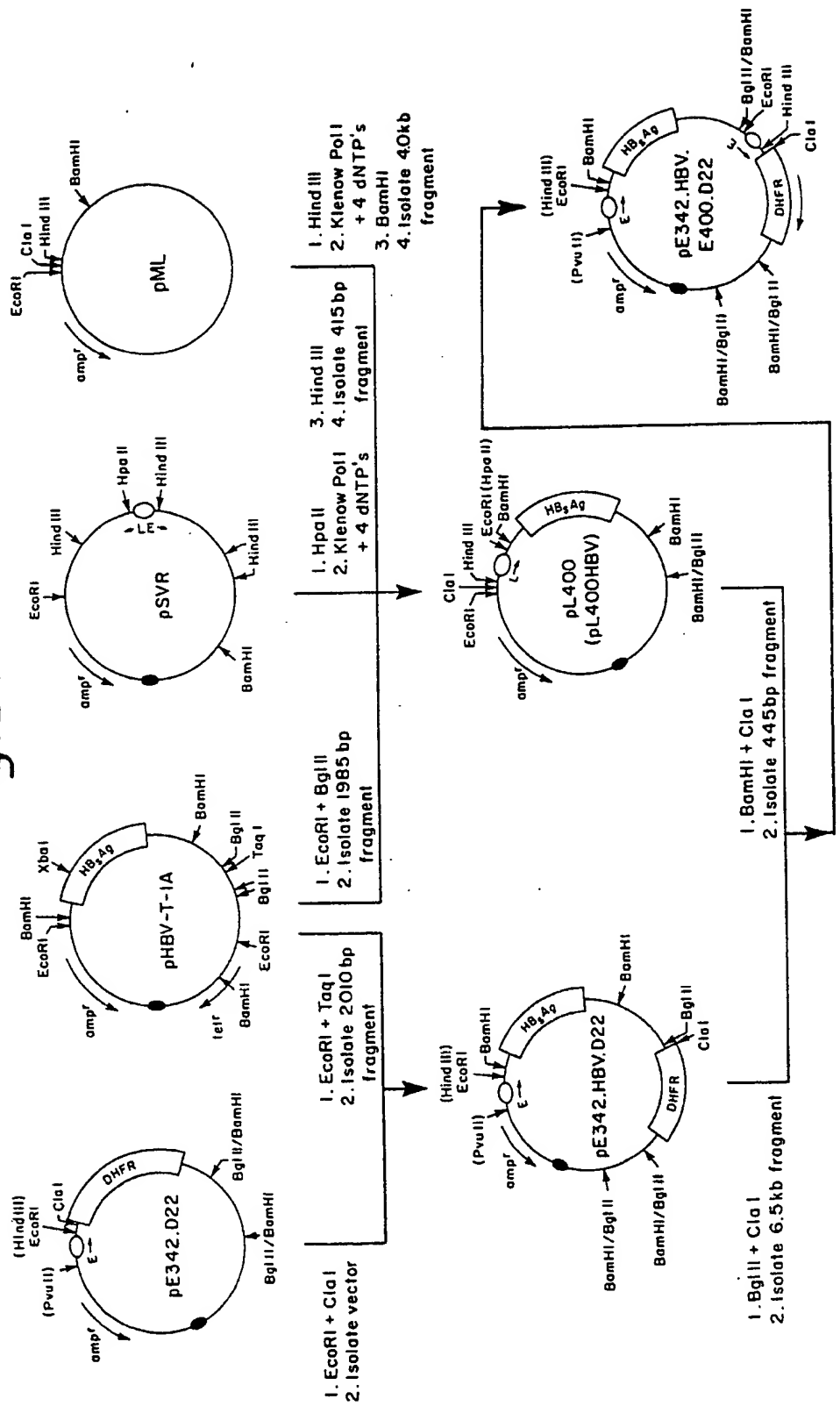

pDHFR-II
DHFR
Pst I
Bgl II
Bgl II
Pst I
1. Pst I
2. Bal 31
3. Bgl II
4. Isolate 660bp fragment
5. Ligate into Hind III / Klenow + Bgl II vector
pDHFR-D22
DHFR
EcoRI
Cla I
Bgl II
Pvu II
ampr
1. EcoRI + Bgl II
2. Isolate 700bp fragment
1. EcoRI + BamHI
2. Isolate 3.4 kb vector
pE342ΔRI
HBsAg
(Hind III)
EcoRI
BamHI
(Pvu II)
E
ampr
BamHI
BamHI/Bgl II
pE342.D22
DHFR
(Hind III)
EcoRI
Cla I
(Pvu II)
E
ampr
Bgl II/BamHI
Bgl II/BamHI

Fig. 2.

Fig. 3.


pML
pSVR
pHBV-T-1A
pE342.D22
pL400 (pL400HBV)
pE342.HBV.D22
pE342.HBV. E400.D22
1. EcoRI + Cla I 2. Isolate vector
1. EcoRI + Taq I 2. Isolate 2010 bp fragment
1. EcoRI + Bgl II 2. Isolate 1985 bp fragment
1. Hpa II 2. Klenow Pol I + 4 dNTP's
3. Hind III 4. Isolate 415 bp fragment
1. Hind III 2. Klenow Pol I + 4 dNTP's 3. BamHI 4. Isolate 4.0kb fragment
1. BamHI + Cla I 2. Isolate 445bp fragment
1. Bgl II + Cla I 2. Isolate 6.5kb fragment